



INSTITUTO POLITÉCNICO NACIONAL

ESCUELA SUPERIOR DE CÓMPUTO

NOMBRES:

PADILLA MATIAS CRISTIAN MICHEL

SAUCILLO GONZÁLEZ JESSE OBED

GRUPO:

4BM1

TRABAJO:

Practica 11

"Arboles de decisión"

MATERIA:

Fundamentos de Inteligencia Artificial

FECHA

8 de diciembre del 2024

INSTITUTO POLITÉCNICO NACIONAL



ESCOM

PRACTICA 9. Árboles de decisión

RESUMEN

Este trabajo se centra en el uso de árboles de decisión utilizando el módulo `sklearn.tree` para clasificación. Se destaca la importancia de transformar variables categóricas a numéricas y se ilustra cómo realizar predicciones con el modelo resultante. Este trabajo no solo se enfoca en la aplicación de árboles de decisión en clasificación, sino que también destaca la importancia de la transformación de variables categóricas, la visualización del modelo y la interpretación de decisiones para maximizar la eficacia de este enfoque en la resolución de problemas clasificatorios.

INTRODUCCIÓN

En el vasto y dinámico campo de la inteligencia artificial, los árboles de decisión se erigen como herramientas fundamentales y versátiles, desempeñando un papel crucial en la resolución de problemas tanto de clasificación como de regresión. Estas estructuras jerárquicas, conceptualmente similares a las decisiones tomadas por seres humanos en situaciones complejas, se destacan por su capacidad para modelar y desentrañar patrones inherentes en conjuntos de datos diversos. Entre los algoritmos más prominentes que facilitan la construcción de estos árboles se encuentra el conocido como CART (Classification and Regression Trees).

El algoritmo CART, acrónimo de su denominación en inglés, se ha consolidado como una metodología robusta y eficaz para la generación de árboles de decisión. Su aplicabilidad abarca diversas áreas, desde la predicción meteorológica hasta la toma de decisiones médicas, ofreciendo soluciones interpretables y de fácil comprensión para problemas complejos. La esencia de este algoritmo reside en su capacidad para dividir iterativamente el conjunto de datos en subconjuntos más homogéneos, guiándose por criterios como la pureza en el caso de la clasificación o la reducción del error en la regresión.

En este contexto, exploraremos la utilidad y aplicaciones prácticas de los árboles de decisión, examinando de manera específica cómo el algoritmo CART contribuye a

la construcción de modelos predictivos y descriptivos. A través de ejemplos concretos, destacaremos la importancia de la transformación de variables categóricas a numéricas y exploraremos cómo estos árboles se convierten en herramientas esenciales para la toma de decisiones automatizada. Desde la conceptualización inicial hasta la aplicación práctica en escenarios del mundo real, este estudio proporcionará una visión integral de los árboles de decisión y su papel esencial en el panorama actual de la inteligencia artificial.

DESARROLLO

Para esta práctica se implementó en Python el algoritmo CART, utilizando el módulo `sklearn.tree` el cual incluye modelos basados en árboles de decisión para clasificación y regresión.

`LabelEncoder` sirve para transformar las variables categóricas a datos numéricos.

```
14 #Adquiere los datos desde un archivo csv usando biblioteca PANDAS
15 dataframe = pd.read_csv('med.csv', encoding='ISO-8859-1')
16
17 #Preprocesamiento (Transformar valores categóricos a numéricos)
18 Edad=LabelEncoder()
19 Sexo=LabelEncoder()
20 PresionSanguinea=LabelEncoder()
21 Colesterol=LabelEncoder()
22 Medicamento=LabelEncoder()
23
24 dataframe['Edad']=Edad.fit_transform(dataframe['Edad'])
25 dataframe['Sexo']=Sexo.fit_transform(dataframe['Sexo'])
26 dataframe['PresionSanguinea']=PresionSanguinea.fit_transform(dataframe['PresionSanguinea'])
27 dataframe['Colesterol']=Colesterol.fit_transform(dataframe['Colesterol'])
28 dataframe['Medicamento']=Medicamento.fit_transform(dataframe['Medicamento'])
29
```

Se preparan los datos y se usa la impureza de Gini como criterio para seleccionar al atributo que mejor divide los datos, además, se realiza la función para visualizar el árbol binario.

```
30 #Prepara los datos
31 features_cols=['Edad','Sexo','PresionSanguinea','Colesterol']
32 X=dataframe[features_cols]
33 y=dataframe.Medicamento
34
35 #Entrenamiento
36 tree = skl.DecisionTreeClassifier(criterion='gini')
37 tree.fit(X,y)
38
39 #Visualización
40 px = 1/plot.rcParams['figure.dpi'] # Pixel in pulgadas
41 fig = plot.figure(figsize=(1000*px,1000*px))
42 _ = plot_tree(tree, feature_names=features_cols, class_names=['A','B'], filled=True)
43
```

Por último, se prueba el modelo y se realiza e imprime la predicción del árbol.

```

45 # Probar el Modelo
46
47 dfprueba = pd.DataFrame()
48 dfprueba[ 'Edad' ] = [1] # Mediana-Edad
49 dfprueba[ 'Sexo' ] = [0] # F
50 dfprueba[ 'PresionSanguinea' ] = [1] # Baja
51 dfprueba[ 'Colesterol' ] = [1] # Normal
52
53
54 prediccion = tree.predict(dfprueba)
55
56 print('Resultado de La prueba')
57 print('*****')
58 print('Con los datos')
59 print(dfprueba)
60 print('\nSe recomienda:')
61 if(prediccion[0]==0):
62     print('El medicamento A')
63 else:
64     print('El medicamento B')
65 print("*****")

```

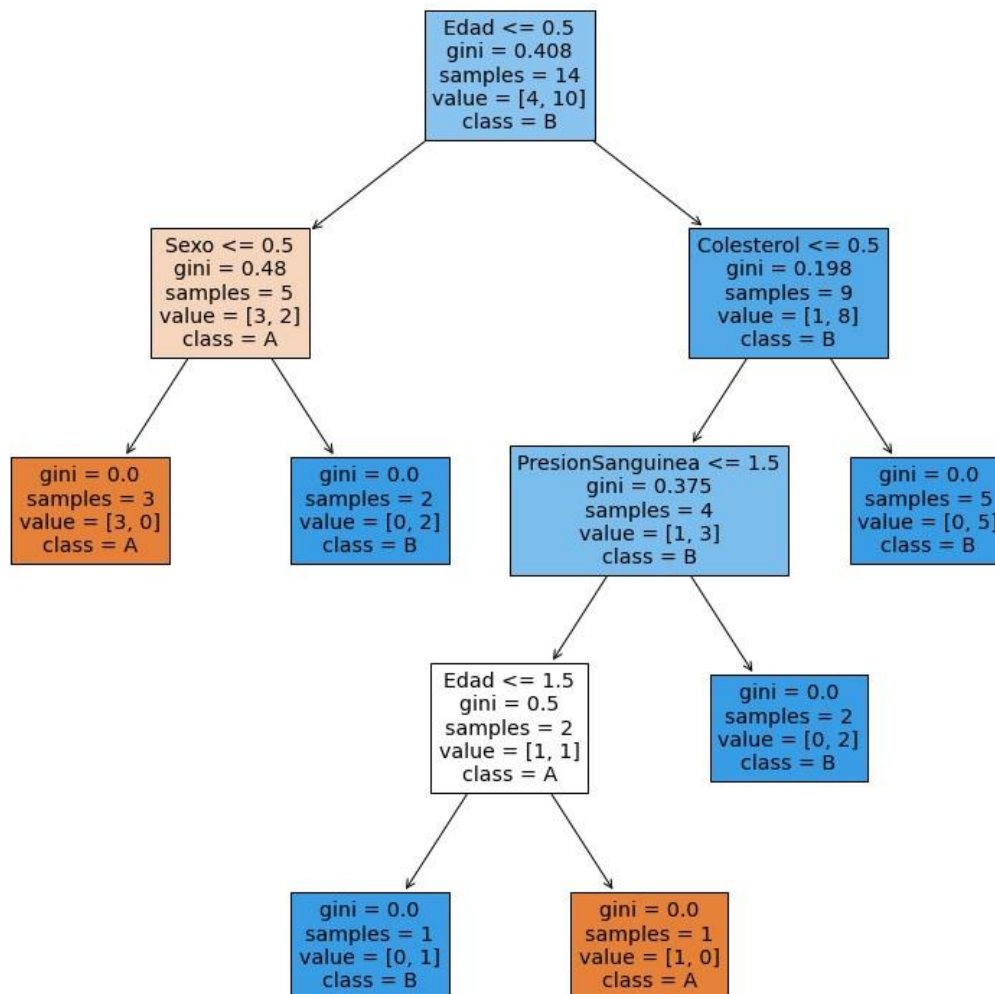
Conjunto de entrada original de datos para el algoritmo

Paciente	Edad	Sexo	PresionSanguinea	Colesterol	Medicamento
p1	Joven	F	Alta	Normal	A
p2	Joven	F	Alta	Alto	A
p3	MedianaEdad	F	Alta	Normal	B
p4	Senior	F	Normal	Normal	B
p5	Senior	M	Baja	Normal	B
p6	Senior	M	Baja	Alto	A
p7	MedianaEdad	M	Baja	Alto	B
p8	Joven	F	Normal	Normal	A
p9	Joven	M	Baja	Normal	B
p10	Senior	M	Normal	Normal	B
p11	Joven	M	Normal	Alto	B
p12	MedianaEdad	F	Normal	Alto	B
p13	MedianaEdad	M	Alta	Normal	B
p14	Senior	F	Normal	Alto	B

Datos transformados:

Edad	Sexo	PresionSanguinea	Colesterol	Medicamento
0	0	0	1	0
0	0	0	0	0
1	0	0	1	1
2	0	2	1	1
2	1	1	1	1
2	1	1	0	0
1	1	1	0	1
0	0	2	1	0
0	1	1	1	1
2	1	2	1	1
0	1	2	0	1
1	0	2	0	1
1	1	0	1	1
2	0	2	0	1

El árbol binario que resultó de aplicar el algoritmo CAR:



Para consultar el árbol, debemos considerar los valores numéricos de cada atributo. Las tablas de equivalencia son:

Para Edad:

Valor	Significado
0	Joven
1	Mediana-Edad
2	Senior

Para sexo:

Valor	Significado
0	F
1	M

Para presión sanguínea:

Valor	Significado
0	Alta
1	Baja
2	Normal

Para colesterol:

Valor	Significado
0	Alto
1	Normal

Para medicamento:

Valor	Significado
0	A
1	B

De acuerdo con el árbol construido ¿qué medicamento se le debe recomendar a una chica de mediana edad, con presión sanguínea baja y colesterol normal?

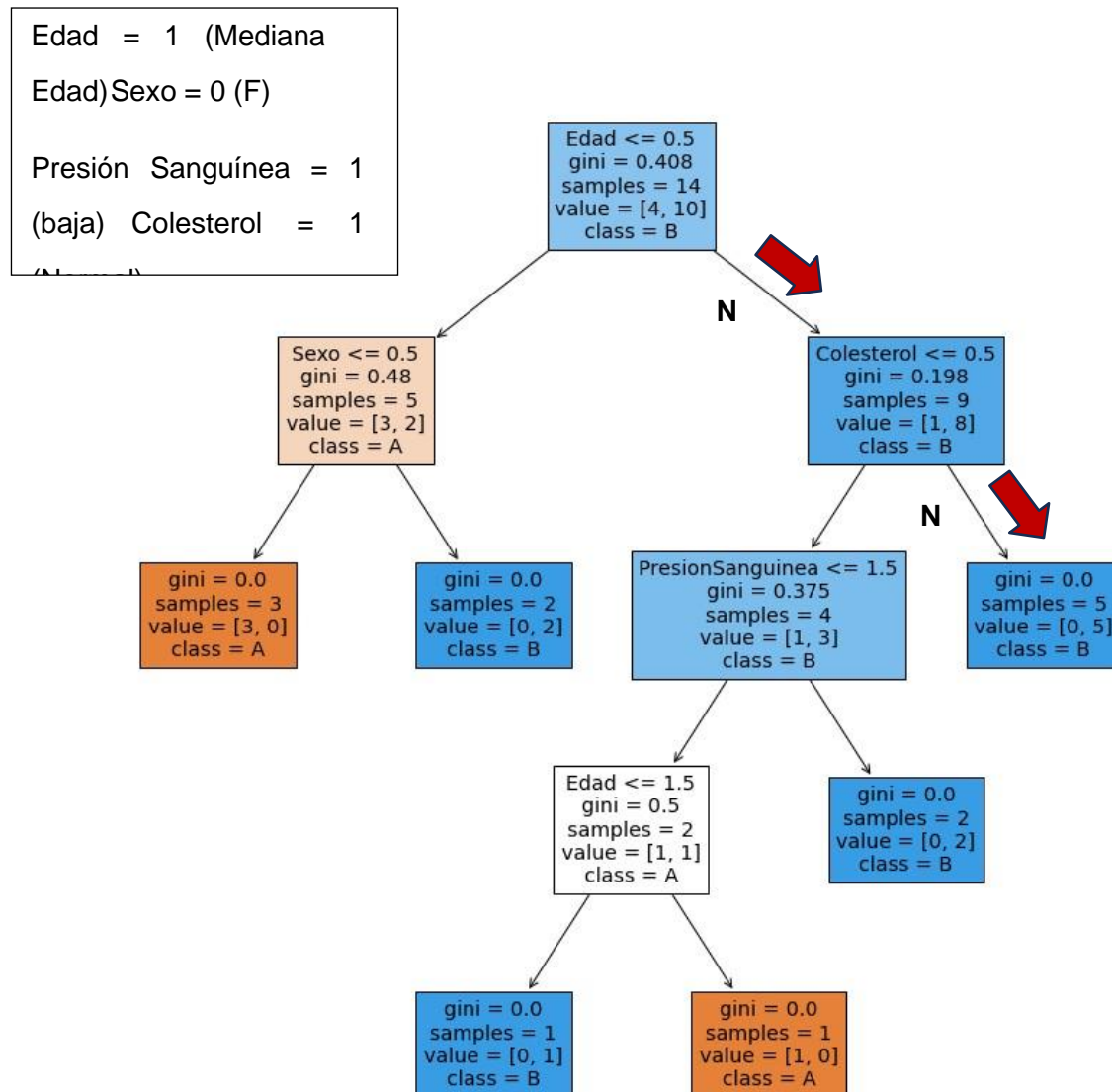
Al pedirle predicciones al modelo también deben hacerse las transformaciones correspondientes. Para realizar esta consulta, y de acuerdo con los datos transformados (numéricos), se considerarán los datos:

```

49 dfprueba[ 'Edad' ] = [1] # Mediana-Edad
50 dfprueba[ 'Sexo' ] = [0] # F
51 dfprueba[ 'PresionSanguinea' ] = [1] # Baja
52 dfprueba[ 'Colesterol' ] = [1] # Normal

```

De acuerdo con el árbol de decisión:



De acuerdo con el árbol de decisión, con las condiciones dadas, el medicamento que se debe recomendar es el B.

El resultado en pantalla:


```

In [1]: runfile('C:/Users/vfrid/Downloads/Práctica 10_F
Downloads/Práctica 10_FIA')
Resultado de la prueba
*****
Con los datos
  Edad  Sexo  PresionSanguinea  Colesterol
0     1     0                 1           1

Se recomienda:
El medicamento B
*****

```

DISCUSIONES DE RESULTADOS

El árbol de decisión proporciona una herramienta visual para entender cómo se toman las decisiones basadas en las variables proporcionadas. En este caso, el modelo utiliza la información de Edad, Sexo, Presión Sanguínea y Colesterol para recomendar un medicamento específico.

Se observan los datos totales con los del árbol, se puede observar que 10 de los 14 datos se evaluaron con dos condiciones, y para los 4 restantes, se necesita evaluar más de dos veces, lo que nos dice que poco más de dos terceras partes de los datos son posibles de concluir con 2 criterios.

En el caso de los datos correspondientes a nuestra prueba, los criterios fueron tales que sólo fue necesario evaluar dos condiciones para concluir el resultado, los cuales son los mínimos necesarios, por lo tanto, fue una evaluación óptima sobre nuestro árbol.

CONCLUSIONES

La aplicación de árboles de decisión en la toma de decisiones como recomendación de medicamentos o situaciones meteorológicas demuestra ser una herramienta valiosa y prometedora en el campo de la inteligencia artificial en la medicina. La combinación de técnicas de modelado de datos y la consideración de factores clínicos específicos permite la generación de recomendaciones personalizadas y respaldadas por datos.

BIBLIOGRAFÍA

TecScience. (2023, 10 septiembre). *Árboles de decisión: creando modelos inteligentes capaces de explicar su respuesta*. <https://tecscience.tec.mx/es/divulgacion-ciencia/arboles-de-decision-creando-modelos-inteligentes-capaces-de-explicar-su-respuesta/>

Árboles de decisión: qué son y cuál es su uso en big data. (2023, 17 abril). UNIR México. <https://mexico.unir.net/ingenieria/noticias/arboles-de-decision/>